PROMOTING OPPORTUNITIES TO LEARN FOR STATISTICIANS

Layla Guyot and Alexander White
The University of Texas at Austin
Texas State University, San Marcos, TX
layla.guyot@austin.utexas.edu

Although the demand for statisticians is growing across many different fields and sectors and more and more students are graduating with degrees in statistics, employers express difficulties in hiring. Indeed, there seems to be a lack of continuity during the transition to the workplace for statisticians. To study this phenomenon, we interviewed junior statisticians who recently experienced the transition as well as mentors who supervised it. A cross-case analysis allowed for an in-depth description of the development of statistical practices in academic settings and at the workplace. Based on our findings and from the perspectives of the participants, we formulate recommendations to promote different practices that are especially challenging to facilitate the transition between education and profession for future statisticians.

THE ROLE OF STATISTICIANS

Statistical skills are increasingly valuable in the job market, and the employment of statisticians in the United States is expected to grow by 35% between 2020 and 2030, which is much faster than the projected 8% growth rate for all occupations (U.S. Bureau of Labor Statistics, 2022). However, the job market offers very few entry-level openings for statisticians, and graduates in statistics are encountering many challenges as they transition to the profession. Although students tend to focus on statistical theory and concepts in academic settings, statisticians perform practices that grow out of experience and are relevant to their discipline and context (Pfannkuch & Wild, 2000).

What Statisticians Do

Practices that statisticians engage with have been considered from different perspectives. Making a list of statistical practices is quite challenging because statisticians are involved in so many different types of activities. Practices were identified at different levels: broadly conceptualized such as *performing data analysis* (Harraway & Barker, 2005) or very specifically observed in one context (Bakker et al., 2008). Cameron (2009) identified four main aspects of statistical practices: formulating a problem to be addressed through statistical means, preparing data (including data collection, organization, and validation), analyzing data with models or other summaries, and presenting results in written, graphical, or other forms. We also consider skills needed to engage in those activities as practices; for example, using statistical software packages or communicating with non-statisticians.

Previous studies (Harraway & Barker, 2005; Hijazi et al., 2019) investigated statistical practices at the workplace, assessing how some specific statistical techniques and research methods were commonly used across many disciplines and types of positions. For example, these studies found that statisticians tended to focus on descriptive statistics more than inferences, modeling, or advanced techniques. To deepen our understanding of the importance of some statistical practices relative to one another, we asked members of the statistical community to rank a list of practices in order of importance (Guyot & White, 2019). We found that translating a real problem into a statistical form and applying statistical methods and techniques appeared to be the most important practices.

Learning in Transition

What students learn in academic settings does not perfectly match what is required of professionals at the workplace. Grosemans et al. (2017) revealed that many challenges occur during the transition across diverse educational backgrounds and disciplines, and a few studies have explored the transition experienced by statisticians in particular. Comparing the perspectives of junior statisticians and their managers, Osman and Ismail (2009) reported that some practices were appropriately developed in academic settings (making graphs or producing descriptive analysis) whereas others needed further development at the workplace (designing databases and using simulation techniques). Indeed, Van der Berg (2017) found that most statisticians agreed that they

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

acquired the appropriate statistical *knowledge* needed at the workplace, although they did not acquire the appropriate statistical *skills*.

The role of mentors at the workplace was highlighted by Hijazi et al. (2019), with many participants describing learning new statistical knowledge at the workplace through mentoring. While junior statisticians reported experiencing teamwork in academic settings, they found integrating larger scale projects challenging at the workplace. According to mentors, junior statisticians were well prepared for programming except in the context of big data and needed further support for written communication skills such as summarizing and organizing ideas.

Although a misalignment between education and profession has been identified for statisticians, the goal of our study was to give the opportunity to statisticians and their mentors to share their experiences and how they overcame the challenges, revealing the mechanism involved in developing statistical practices.

RESEARCH DESIGN

We first met members of the statistical community at seven conferences worldwide where attendees were invited to engage in a sorting activity (Guyot & White, 2019) to initiate their reflection about the role of statisticians. After recruiting 154 attendees, we purposefully sampled eight junior statisticians as well as two mentors who have accompanied junior statisticians during the transition to conduct interviews. A cross-case analysis allows for an in-depth description of the phenomenon of the transition and the development of statistical practices.

Participants and Data Collection

We invited eligible participants who maximized the variation between fields of applications and types of degrees for junior statisticians. All eight junior statisticians we interviewed earned a master's degree but within various specialties: applied statistics, biostatistics, or public health. They worked in diverse sectors (health care, government, policy, education, economics), and they had between one and three years of experience. The two mentors included in our study had over eight years of experience supervising junior statisticians during their transition in the fields of policy and health care. One of the mentors was also teaching a course in academia.

We conducted one-on-one online interviews between May and October 2019. The interviews lasted between 45 and 115 minutes, allowing for in-depth descriptions of the unique experiences of each participant. Participants' experiences offered multiple cases for learning practices during the transition to the workplace. Each interview had a tailored interview protocol and was audio recorded and transcribed.

Framework and Data Analysis

We used the theoretical framework of activity systems (Konkola et al., 2007; Vygotsky, 1978) and boundary crossing (Akkerman & Bakker, 2011) to conceptualize the transition (Figure 1). We interpreted the transition as an interaction between two activity systems, academic settings and the workplace, whose goals are to develop statistical practices, with each system involving different elements.

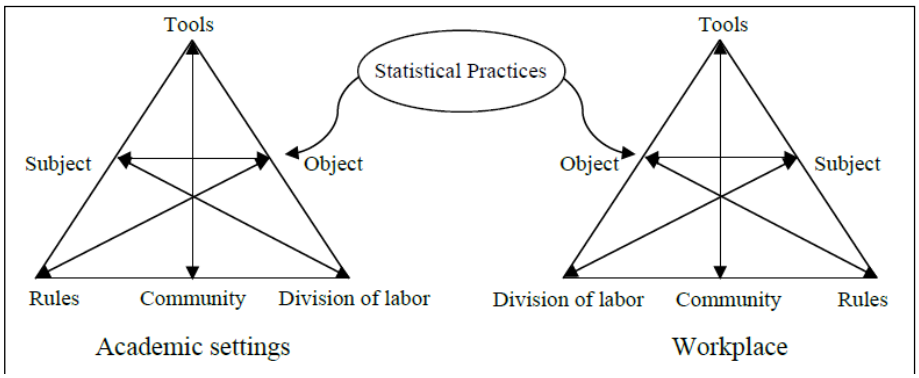

Figure 1. Interaction between activity systems

An activity occurs within a system and is mediated by tools and regulated by rules, with the division of labor ensuring the distribution of tasks within a community. While transitioning between the two systems, statisticians encountered boundaries arising from the lack of development of some practices. For example, in academic settings, a student (*subject*) develops statistical practices (*object*) using curricula materials (*tools*). The *community* of students, teachers, and advisors work together to ensure that the student meets the degree requirements (*rules*). The teacher defines the responsibilities (*division of labor*) of each participant. At the workplace, a statistician (*subject*) develops statistical practices (*object*) using statistical software (*tools*). The collaboration among statisticians, managers, and clients forms a community, following regulations (*rules*) and a specific organization of the different tasks (*division of labor*).

The analysis was conducted in three main steps. First, we coded the transcripts and created categories of statistical practices, refining these categories constantly. We recognized differences and commonalities between the meanings of the practices across participants and identified main categories of statistical practices. For example, one of the main categories was related to programming practices. For each statistical practice, we looked for quotes illustrating each element in the framework: tools, rules, division of labor, and community; and for each activity system: academic settings and the workplace. In the quotes that mentioned tools for programming in academic settings, we found that some participants took computer science courses that fostered the logic of programming. Finally, we compared the differences and commonalities across participants between the two activity systems to identify boundaries in the tools, rules, division of labor, and community. We used the levels described by Grosemans et al. (2017) to categorize the transition as continuous, detailing, or discontinuous. For example, considering the practice of programming, participants who took computer science courses in academic settings experienced a continuous transition while participants who did not take these types of courses had to further develop (*detail*) their knowledge. No participant considered their transition for programming practices as discontinuous because they all learned how to use at least one programming language or some software to perform statistics.

RESULTS

The cross-case analysis revealed eight categories of statistical practices (Table 1). We identified the element that mostly created boundaries for each practice and reported how many participants experienced a discontinuous, detailing, or continuous transition.

Table 1. Transition of statistical practices

	Boundaries	Discontinuous	Detailing	Continuous
Data Management	Rules	4	2	4
Documentation	Rules	4	0	1
Collaboration	Community	4	3	3
Communicating	Community	3	5	2
Time Management	Rules	3	5	0
Design	Division	2	5	3
Analysis	Division	0	4	6
Programming	Rules	0	3	7

Challenges Encountered

Most boundaries encountered during the transition emerged from differences in *rules* between academic settings and the workplace and had an impact across several practices. Considering the practice of data management, dealing with messy data created boundaries as “almost always in school, they just kind of give you these perfect data sets. … But it's really unrealistic once you get to the real world.” The fact that decision making relies on interpretations of statistics at the workplace

also introduced new rules because interpretations in academic settings were often decontextualized and the difference between a correct and incorrect answer had limited impact. As stated by one of the mentors: "in professional life, I really have to figure out what it actually means because potentially millions and millions of dollars are depending on my decision as opposed to the difference between an A and an A minus." Other boundaries included time constraints for the practices of programming, data management, and analysis, weighing tradeoffs between the best method *theoretically* and the method that *practically* allows the statistician to complete a task in a timely manner.

Recurring boundaries in the *division of labor* concerned the ability to choose the appropriate methodology, analysis technique, or program to analyze data. Junior statisticians often had no experience making such decisions when they entered the workplace because professors were holding the division of labor in academic settings: the structure of courses was usually "based off of a technique and not really based off of understanding the data and then trying to pick the technique to fit the data."

Differences in the *community* also introduced boundaries. At the workplace, participants had to interact with non-statisticians, which is very different from "graduate school where everyone's studying the same thing." Indeed, "communicating with [non-statisticians] what results are from a statistical point of view and what the implications are, has been a lot more challenging."

Transition of Practices

Participants generally experienced continuity with *Programming* and *Analysis*, whereas the categories of *Collaboration*, *Data management*, and *Documentation* had the highest counts for discontinuity. Comparing individual experiences, we identified elements that promoted continuity and others that caused discontinuity.

Computer science courses in undergraduate programs were found to be beneficial to develop programming practices, providing a solid foundation to think logically, and being able to switch between different languages or software. One of the junior statisticians concluded that "the most useful practice [they] learned [in academic settings] was computer programming. [They] have to pull [their] own data and learning SQL was vital to do this. Learning to program in multiple languages makes you more marketable as a statistician."

Indeed, having acquired various programming tools in the academic settings supported the development of other practices at the workplace. In particular, one of our participants highlights how learning to use Jupyter Notebooks sustained documentation practices: "you could output your results right in line with the code like you just hand someone your lab notebook. … you could explain what your code is doing and how you got that code, that graph." Another participant actually experienced discontinuity for documentation as they struggled to learn how to use GitHub at the workplace: "it took me literally like a year and a half to understand what the code was trying to do, after begging people to explain it to me."

Detailing experiences for collaboration practices emphasized some examples of elements present in academic settings that eased the transition. Overall, participants suggested consulting courses or laboratories such as offering statistical support within the university or to local organizations. These opportunities were only found to be authentic if they involved domain experts or real clients. One participant also mentioned that their experience in customer service helped develop skills required to understand clients' needs. Similarly, junior statisticians reported needs for more support towards communication practices, by providing the opportunity to rehearse and emphasizing the importance of interpretations for decision making rather than technicalities. Two participants had the opportunity to take classes where "students learn how to communicate, and they talked a lot about body language and things like that. And it would be good to have that. You know either in undergraduate courses or even in graduate courses."

Mentoring

We found that the role of the mentor at the workplace is crucial for managing the expectations and priorities during the transition. Some workplaces even offered a mentoring program by assigning a mentor to junior statisticians, which supported learning for various practices. The importance of the role of mentors was particularly uncovered with the experience of one participant who was first isolated by being the only statistician present at the workplace and had to go "back to

my professors once or twice" to get help. Most of the practices for this participant were qualified as discontinuous until a senior statistician who was working remotely decided to act as a mentor.

Participants especially benefitted from a mentor's expertise when developing practices involved in the design of a statistical study because they did not feel they had enough experience or credibility. Mentors provided opportunities both in academic settings and at the workplace to support learning. They recommended that students should be able to practice how to design their own studies in academic settings, yet they should not only pursue their own interests.

Mentors also have the responsibilities to monitor the workload of junior statisticians so that they do not get overwhelmed by having too many projects or tasks assigned to them. In addition, mentors realize how essential their role is in the transition if they had previously received mentorship training.

OPPORTUNITIES TO LEARN

The aim of our study was to provide insights into the mechanisms involved in developing statistical practices while in transition to the workplace. We formulate implications for easing the transition of statistical practices at three levels: in academic settings, at the workplace, and for the statistical community in general.

Implications in Academic Settings

Currently, there are very few courses that promote data management practices and not enough textbooks dealing with related topics. We noticed on Google scholar that studies referring to "teaching data management" were very rare and mainly concentrated on the fields of computer science or library science, with most of the research being published after 2010. Our findings reinforced the need for such courses in the specific field of statistics, emphasizing how to use the appropriate tools and methods to clean messy data, handle missing data, or extract data from databases stored remotely. However, the emergence of data science programs in the past few years has impacted how these topics are included.

Getting experience with the components of attitude and relationship in academic settings would have facilitated the transition for our participants because they were particularly challenged by the attitude to develop while working with domain experts, being diplomatic and understanding, yet not condescending.

To raise awareness about a statistician's responsibilities, participants suggested organizing a seminar before graduation to present reasonable expectations for junior statisticians at the workplace and how to find resources to keep learning at the workplace. Efforts to invite statisticians from the workplace to share their experience in academic settings through panel discussions and alumni reunions were also mentioned.

Implications at the Workplace

Our findings showed that junior statisticians need to keep learning at the workplace across all practices, but this need was particularly imperative for programming. Participants mostly mentioned having to teach themselves new programming tools and referred to examples of workshops, forums, or interest groups organized onsite at the workplace that sustained learning. Most participants were able to go to conferences and mentioned learning about communication, keeping up to date with the newest statistical techniques, or networking. They also emphasized the importance of seeking the help of a mentor and opportunities for professional development. Hence, employers should create mentoring programs, help junior statisticians identify mentors, and provide support for attending conferences and workshops.

Implications for the Statistical Community

Many of our participants shared that they studied statistics by *happenstance* because they were not aware of degrees in statistics or what they could do with such a degree when they started their educational journey. As a matter of fact, the lack of visibility for the profession may contribute to the shortage of statisticians on the job market. The statistical community needs to promote the role of statisticians, with interventions early in education, by publishing in different journals, and organizing interdisciplinary conferences.

CONCLUSION

The aims of teaching statistics should not only focus on the knowledge of statistical theory, methods, or applications but also develop many skills desirable for a statistician. We conducted this study to gain a deeper understanding of the elements that can facilitate the transition of statistical practices between academic settings and the workplace. Experiences were considered more continuous if mentors accompanied statisticians in the transition at the workplace, if practices were incorporated across the curriculum in academic settings, and if students were given the authority to choose appropriate tools and methods.

Investigating the transition between academic settings and the workplace for statisticians is challenging. Participants in our study experienced various contexts and because they identified common practices across their differences, the findings yield significant contributions. However, we recommend narrowing the profile of participants to those that share the same context (graduating from the same program or working at the same workplace), to better illustrate the relationship between the levels of transition to the workplace and the background in academic settings across individuals.

REFERENCES

Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*(2), 132–169. https://doi.org/10.3102%2F0034654311404435

Bakker, A., Kent, P., Derry, J., Noss, R., & Hoyles, C. (2008). Statistical inference at work: Statistical process control as an example. *Statistics Education Research Journal, 7*(2), 130–145. http://doi.org/10.52041/serj.v7i2.473

Cameron, M. (2009). Training statisticians for a research organisation. In *Proceedings of the 57th Session of the International Statistical Institute. Statistics: Our past, present & future.* International Statistical Institute.

Grosemans, I., Coertjens, L., & Kyndt, E. (2017). Exploring learning and fit in the transition from higher education to the labour marker: A systematic review. *Educational Research Review, 21*, 67–84. https://doi.org/10.1016/j.edurev.2017.03.001

Guyot, L. & White, A. (2019). Promoting an authentic experience of statistical practices in statistics education. In S. Otten, A. G. Candela, Z. de Araujo, C. Haines, & C. Munter (Eds.), *Proceedings of the forty-first annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 421–425). University of Missouri.

Harraway, J. A., & Barker, R. J. (2005). Statistics in the workplace: A survey of use by recent graduates with higher degrees. *Statistics Education Research Journal*, *4*(2), 43–58. https://doi.org/10.52041/serj.v4i2.514

Hijazi, R., Saeed, R., & Alfaki, I. (2019). Role of statisticians in building the UAE knowledge economy. *Electronic Journal of Applied Statistical Analysis, 12*(1), 303–319. http://doi.org/10.1285/i20705948v12n1p303

Konkola, R., Tuomi-Gröhn, T., Lambert, P., & Ludvigsen, S. (2007). Promoting learning and transfer between school and workplace. *Journal of Education and Work, 20*(3), 211–228. https://doi.org/10.1080/13639080701464483

Osman, M. I., & Ismail, M. A. (2009). What do statisticians working in policy research need from statistical education? In *Next steps in statistics education: Proceedings of the Satellite Conference of the International Association for Statistical Education.* IASE. https://iase-web.org/documents/papers/sat2009/9_4.pdf?1402524995

Pfannkuch, M., & Wild, C. J. (2000). Statistical thinking and statistical practice: Themes gleaned from professional statisticians. *Statistical Science, 15*(2), 132–152.

U. S. Bureau of Labor Statistics. (2022). *Occupational outlook handbook*. https://www.bls.gov/ooh/

Van der Berg, G., H. (2017). *A framework to integrate the formal learning with the informal workplace learning of statisticians in a developmental state* [Doctoral dissertation, University of South Africa]. Unisa ETD. http://hdl.handle.net/10500/22638

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.